Finding Correctly Folded Active RNA Motifs

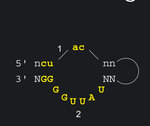
R Knight[1] H De Sterck[2] R Markel[3] C Lozupone[1] S Changayil[1] I Majerfeld[1] M Yarus[1]

[1]MCD Biology, [2]Applied Math, University of Colorado; [3]NOAA: Boulder, CO, 80309, USA

abstract

RNA motifs, also called active sites, consist of a few functional modules separated by unimportant spacer.

We have been investigating two well-characterized motifs:



The Isoleucine Aptamer

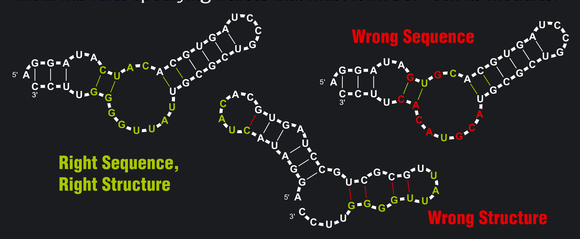
The Hammerhead Ribozyme

Through a combination of analytical calculations, supercomputer simulations, and laboratory experiments, we find that:

1. Functional RNA motifs may be far more common than expected, perhaps allowing a zeptomole RNA world.
2. The length and composition of randomized RNA pools have dramatic effects on folding.

methods

Motifs are defined by inspection of sequence alignments. A motif consists of a set of modules. Each module has a characteristic sequence and structure (where the structure is list of the paired bases. Also, a motif has rules specifying helices that must form between its modules.

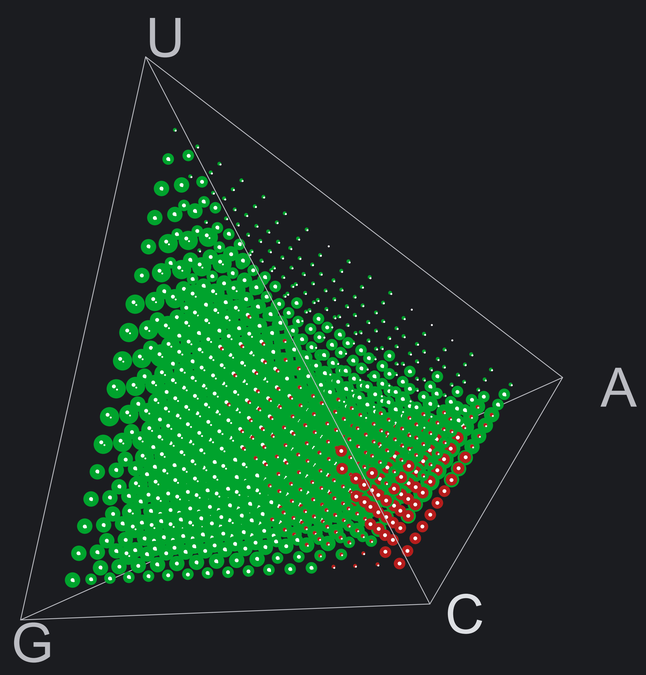


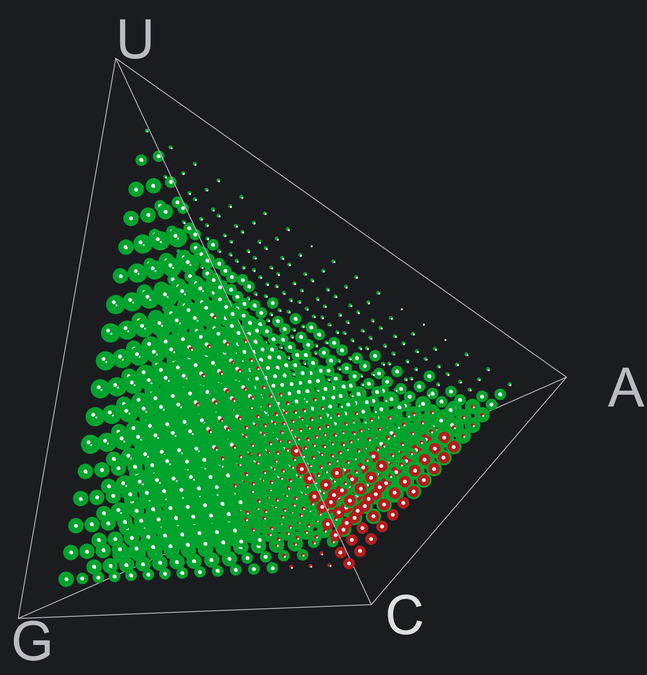
Frequencies of motifs in random sequences can be calculated analytically, giving Pr(sequence). We must find the probability that at least one of the (s+1)!/(s+1-m)!m! ways of placing m modules in s nucelotides of spacer happens to match all the parts of the motif in order (see calculations section).

Folding was characterized using the Vienna RNA package v. 1.4 on the Linux Platinum supercomputer at NCSA. We generated sequences for each module, inserted them into longer random sequences, and asked how often the correct base pairs formed: Pr(structure | sequence). To get Pr(structure & sequence), we multiplied this folding frequency by the frequency of the sequence.

SELEX experiments were performed by selecting aptamers against Ile-sepharose, eluting with 10 mM free L-Ile in 50 mM HEPES, 300 mM NaCl, 7.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 7.0. Aptamers were selected from pools with 16, 22, 26, 35, 50, and 75 randomized nucleotides over 6-11 rounds of selection, starting with 2-9 x 10^{14} template DNA molecules. Results were compared with an earlier selection for Ile binders using a similar protocol and a random region of 50 nucleotides.

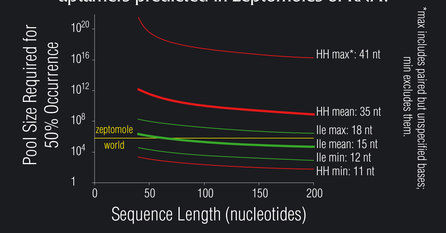
results



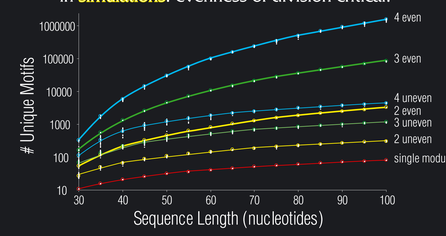


Stereo pair showing folding frequencies of Isoleucine aptamers and Hammerhead ribozymes at 5% intervals in the space of possible RNA compositions. Higher-resolution calculations at 2% steps took 600 CPU hours total spread over 260 processors on the Linux Platinum cluster at NCSA, stepping over 18,424 compositions with 73,696,000 random sequences (50 nucleotides per sequence). Largest points correspond to 30% correct folding; smallest points to 0.1% correct. Isoleucine aptamers fold best in sequences biased towards purines (A+G), and typically folds better than the hammerhead site. The Hammerhead site folds best with amino bases (A+C), probably because of reduced competition.

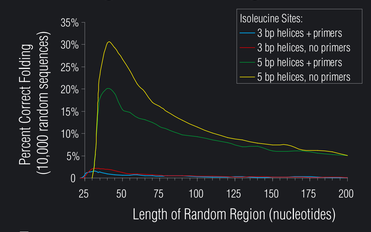
Motif Frequencies: functional sites such as Ile aptamers predicted in zeptomoles of RNA?



Calculations exactly predict the motif sampling in **simulations**: evenness of division critical.

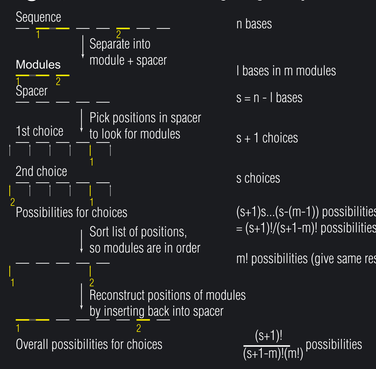


Length poisoning reduces the probability of correct folding as the random region increases.

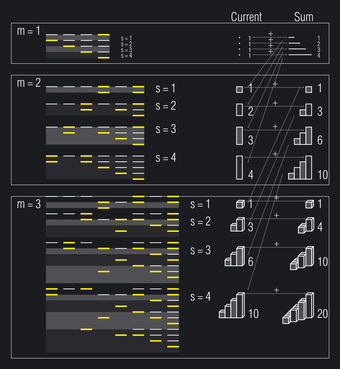


calculations

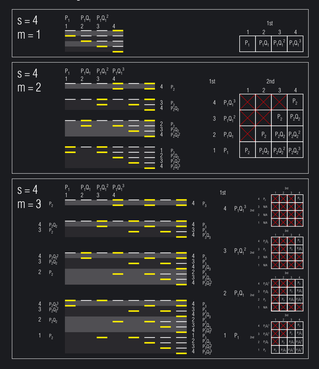
Finding several modules in longer sequences is much easier than finding a single conserved region, since there are many ways to place them.



The amount of spacer s determines the size of the calculation, while the number of modules m determines the dimension.



Since the same sequence is resampled, we calculate the probability recursively in terms of smaller problems in lower dimensions.



conclusions

Modularity makes functional RNA motifs far more frequent than would otherwise be expected, possibly allowing a zeptomole RNA world (functions in less than a million random molecules).

Selected RNA motifs should be as short, modular, and evenly divided as possible.

Primers have a marked influence on folding: it may be possible to select primer sets that avoid interfering with known active structures.

Length and composition of random pools have large effects (several orders of magnitude) on the intrinsic frequencies of functional motifs and on the probability of folding correctly. Length effects on folding in simulations match experiments.

References: Knight & Yarus 2003 RNA 9:218-30; Yarus & Knight 2003 in 'The Genetic Code and the Origin of Life', Landes Bioscience; Lozupone et al. in prep; Majerfeld & Yarus 1998 RNA 4:471-8; Schultes et al. 1997 RNA 3:792-806 Sabeti et al. 1997 Chem Biol 4:767-74; Hofacker et al. 1994 Monatsh Chem 125:167-188; Salehi-Ashtiani & Szostak 2001 Nature 414:82-4.